APPLICATION OF COLLECTIVE RISK THEORY TO ESTIMATE VARIABILITY IN LOSS RESERVES

ROGER M. HAYNE

Abstract

The intent of this paper is to present an introduction to Collective Risk Theory for the first time reader and considerations in applying that theory to estimate variability in loss reserves. It begins with a brief introduction to the basic concepts of Collective Risk Theory along with a survey of some of the techniques developed to date to estimate the aggregate distribution of losses. With this framework, descriptions of some applications to loss reserves are discussed, with attention paid to the assumptions inherent in those methods and some problems that arise in applying this theory to reserves. Of note are questions that are not directly addressed by this model; in particular, parameter uncertainty. Included are references which, it is hoped, will lead the interested reader further into the applications to date.

1. INTRODUCTION

The question of the amount of variability inherent in loss reserve estimates has gained more notice in recent years. In fact, Principles 3 and 4 of the Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves [1] state,

> "3. The uncertainty inherent in the estimation of required provisions for unpaid losses or loss adjustment expenses implies that a range of reserves can be actuarially sound. The true value of the liability for losses or loss adjustment expenses at any accounting date can be known only when all attendant claims have been settled.
>
> 4. The most appropriate reserve within a range of actuarially sound estimates depends on both the relative likelihood of estimates within the range and the financial reporting context in which the reserve will be presented."

Quantification of the variability in reserve estimates will thus be useful in the determination mentioned in Principle 4. In addition, knowledge of the statistical distribution of reserves is also useful in discussing the impact of reserve discounting on insurer capacity and solidity. One author has already cited this as a favorable result of discounting in that discounting of reserves would "increase the statutory capacity of the insurance industry. Statutory surplus would increase as loss reserve liabilities were reduced [2]." However, simply discounting reserves will not necessarily increase financial strength or capacity. Rather, a better measure of that capacity is probably the ability of surplus to protect solvency. Without knowledge of the variability of the reserve estimates, the assessment of the strength of a company at a given level of surplus, and hence capacity, probably cannot be made accurately.

There are several techniques which are available to assess the financial solidity of a given amount of surplus. Methods that have been advanced for this purpose include "confidence limit" approaches, Ruin Theory, and Utility Theory, along with a rather comprehensive model of the operations of an insurer (see [3] and [4] for this latter application). In each case, however, their application requires an estimate of the statistical distribution of the reserves.

The intent of this paper is to discuss the framework of Collective Risk Theory as one approach that can be used to estimate the statistical distribution of reserves. No prior exposure to Collective Risk Theory is assumed; however, it is hoped that the references will provide a good starting place for the reader who wants to pursue this subject further.

2. THE COLLECTIVE RISK MODEL

The basic collective risk model approaches the question of the distribution of total reserves by modeling the claim process faced by an insurer. It considers the interaction between the distribution of the number of claims and the distribution(s) of the individual claims by calculating loss (or reserve) T as the sum

$$T = X_1 + X_2 + \cdots + X_N, \tag{2.1}$$

where the number of claims N is randomly selected, and each of the claims $X_1, X_2, \ldots, X_N$ is randomly selected from claim size distribution(s).

There is a significant amount of literature which addresses this model and its applications to casualty insurance. The primary source is probably the text by Beard, Pentikäinen and Pesonen [5]. Other complete texts dealing with Collective Risk Theory and its applications are those by Borch [6], Bühlmann [7] and Seal [8]. The papers by Borch [9] and Pentikäinen [10] also consider this model from a fairly broad point of view.

There are some useful properties of the distribution T under rather broad assumptions. In particular, if

1. The number of claims N has moments

   $\nu = \mathrm{E}(N)$

   $\nu_i = \mathrm{E}[(N - \nu)^i]$ for $i = 2, 3,$ and 4;

2. All claims are drawn from the same population with moments

   $x = \mathrm{E}(X)$

   $x_i = \mathrm{E}[(X - x)^i]$ for $i = 2, 3,$ and 4; and

3. All claims X and the number of claims N are all independent, then the first four moments of the random variable T exist and are given by

$$\mathrm{E}(T) = \nu x \tag{2.2}$$

$$\mathrm{E}[\{T - \mathrm{E}(T)\}^2] = x_2\nu + x^2\nu_2 \tag{2.3}$$

$$\mathrm{E}[\{T - \mathrm{E}(T)\}^3] = x_3\nu + 3x_2 x\nu_2 + x^3\nu_3 \tag{2.4}$$

$$\mathrm{E}[\{T - \mathrm{E}(T)\}^4] = x_4\nu + 3x_2^2(\nu_2 - \nu + \nu^2) + 4xx_3\nu_2 + 6x^2x_2(\nu_3 + \nu\nu_2) + x^4\nu_4. \tag{2.5}$$

Comparable formulae for higher moments can also be derived if the corresponding moments of the claim count and size distributions exist. The paper by Mayerson, Jones and Bowers [11] gives a derivation of these formulae.

These facts can and should be used to test the reasonableness of any approximation to the distribution of T. In fact, one of the methods used to approximate that distribution relies on these relationships.

3. APPROXIMATIONS OF THE DISTRIBUTION OF T

There have been many approaches used in estimating the distribution of T, given distributions for the number of claims N and the size of those claims. These methods can be broadly grouped into 3 classes:

1. Monte Carlo Simulation,
2. Approximate Distributions, and
3. Analytic Approximation.

Monte Carlo Simulation

Probably the most flexible of these approaches is that of Monte Carlo Simulation. The idea is simple and directly follows the basic Collective Risk Model above. Simply stated, the Monte Carlo Simulation algorithm is composed of five steps:

1. Randomly select the number of claims N from the claim count distribution.
2. Randomly select N claims, $X_1, X_2, \ldots, X_N$, from the claim size distribution.
3. Calculate one observation from the distribution of T by the sum $X_1 + X_2 + \cdots + X_N$.
4. Repeat steps 1 through 3 "several" times.
5. Estimate the distribution of T using the points generated in this manner.

Conceptually, there is no limit on the form of the claim count or size distributions used in Monte Carlo Simulation. They can both be discrete or the claim size distribution can be continuous. Simulation with deductibles and/or per claim loss limitations can also be easily handled in this framework. In addition, the combination of several lines of insurance or accident years can also be accommodated without much difficulty.

There are, however, prices to pay. First, the answer to how many is "several" in step 4 is not clear. Often a significant number of simulations must be run to obtain a clear enough picture of the distribution of T to be useful in applications. One technique, though admittedly "brute force," is to compare the results of two sets of simulations, say each of 1,000 trials. If the resulting distributions are "close enough" for the task at hand, the combined distribution could be used as an approximation. If, however, they differ significantly, more trials may be indicated. The moments of the simulated distributions should be compared to the theoretically expected moments in (2.2) through (2.5) to see if the simulation is sufficiently close.

Another practical consideration is how to simulate the random selections from the claim count and claim size distributions. Care should be taken as to the representations of the distributions. Finite distributions, such as those based solely on empirical data, implicitly have upper bounds. Thus, unless those upper bounds are to be explicitly considered in the model, some of the variability inherent in the underlying distribution may be lost.

One solution to this difficulty could be to use analytic distributions, such as the lognormal or Pareto, to estimate the distributions in the "tail." In this way the empirical data could be used, and yet some of the potentially unlimited nature of some risks can be captured.

The process used to make the random selections from the claim size and count distributions may not be obvious. Most computer software packages do provide "random" number generators which correspond to a uniform distribution. In addition, there are algorithms which allow for selections from other distributions, either directly or from selections from the uniform distribution. The very useful text *A Guide to Simulation* by Bratley, Fox and Schrage [12] includes some of these algorithms for a number of statistical distributions. That text also includes listings of computer programs to perform those calculations.

One final consideration regarding Monte Carlo Simulation is the cost in computer time. Factors influencing this time include the complexity of the model used, the expected number of claims $E(N)$, the degree of accuracy required, and the amount of dispersion in the claim size distribution. Simulations involving a great number of expected claims will of course take longer to run. Not immediately obvious, though, is the fact that if the claim size distribution is dispersed (a large standard deviation as compared to the mean), there will generally be a greater number of simulations necessary to achieve a desired level of accuracy than if the claim size distribution is less dispersed.

Approximate Distributions

Another method used to estimate the distribution of T involves assuming a statistical distribution and then using the "known" moments of T to select the parameters of that distribution. Probably falling into this category is the Normal–Power, or NP Approximation. This approach is described in Beard, Pentikäinen and Pesonen [5] and used by Mayerson, Jones and Bowers [11] and by Patrik and John [13]. Although relatively easy to apply, it does not seem to be sufficiently skewed for many casualty applications. However, caution should be taken in applying this approach. It can easily yield misleading results, or even nonsense, if misapplied, especially if the variable to be approximated differs markedly from the normal distribution.

This approach considers a transformation of the variable T which is hoped to be approximately normal. Although the transformation can be carried out to include several moments of the distribution of T, the application in [11] stops at the third moment with the formula:

$$t_0 = m_1 + m_2 z_0 + m_3(z_0^2 - 1)/6 + m_4(z_0^3 - 3z_0)/24 - m_3^2(2z_0^3 - 5z_0)/36, \quad (3.1)$$

where z_0 represents the $100e$ percentile of a standard normal distribution and t_0 represents the approximate $100e$ percentile of the distribution of T. Here

$$m_1 = \mathrm{E}(T)$$
$$m_2^2 = \mathrm{E}[\{T - \mathrm{E}(T)\}^2]$$
$$m_3 = \mathrm{E}[\{T - \mathrm{E}(T)\}^3]/m_2^3$$
$$m_4 = \mathrm{E}[\{T - \mathrm{E}(T)\}^4]/m_2^4 - 3$$

Using formulae (2.2) through (2.5), the various moments of T can be found from those of the claim count and size distributions. The various percentiles of the aggregate distribution can then be approximated.

A similar approach is followed by Venter in [14]. In that paper, transformations of the Beta and Gamma distributions are suggested as forms for the distribution of aggregate losses. The Gamma distribution is also suggested by Beard, Pentikäinen and Pesonen, [5, page 121]. Again, matching of moments is used to estimate the parameters of the distribution. Pentikäinen [15], Lau [16] and Philbrick [17] also present approaches based on distribution fitting.

The benefit of this approach is its relative simplicity and, once the moments are calculated, the ease with which the percentiles of the aggregate distribution can be approximated. It does require, however, that the form of the distribution be assumed and there are no readily available tests of how well the distribution used fits the actual distribution of T.

Analytic Approximation

A third category of approximations of the distribution of T attempts to analytically calculate that distribution. This approach generally looks at the distribution of T as the sum

$$F(t) = \sum_{n=0}^{\infty} \mathrm{P}(N = n)\, F_n(t), \tag{3.2}$$

where $\mathrm{P}(N = n)$ is the probability of n claims and $F_n(t)$ is the probability that the sum of n claims will be less than t. The functions $F_n(t)$ can then be calculated in terms of the probability density function of the individual claim size distribution. In the discrete case, for example, if $F(x)$ is given by

$$F(100) = 0.60$$
$$F(300) = 1.00,$$

then $F_2(x)$ will be given by

$$F_2(200) = 0.36$$
$$F_2(400) = 0.84$$
$$F_2(600) = 1.00.$$

Since there are only two outcomes of the original distribution, a loss of 100 with probability .6 and a loss of 300 with probability .4, the only possible outcomes for the sum are 200 (two losses at 100 each), 400 (one loss at 100 and one at 300), and 600 (two losses at 300 each). The resulting distribution is called the convolution of the probability density function (p.d.f.) underlying F with itself. More generally, in the continuous case, if $f(x)$ and $g(y)$ are p.d.f.'s for independent random variables X and Y, then the sum $Z = X + Y$ has the p.d.f. given by

$$(f^*g)(z) = \int_{-\infty}^{\infty} f(x)g(z-x)dx, \tag{3.3}$$

which is called the convolution of f and g. Similar to multiplication define f^{*n} iteratively by

$$f^{*0} = 1$$
$$f^{*n} = f^* f^{*(n-1)} \text{ for } n = 1, 2, \ldots$$

Then $F_n(x)$ can be written in terms of f^{*n} as

$$F_n(x) = \int_{-\infty}^{x} f^{*n}(z)dz. \tag{3.4}$$

If now the p.d.f. of the claim size distribution is $f(x)$, then, combining (3.2) and (3.4), the p.d.f. underlying the distribution of T can be written as

$$h(t) = \sum_{n=0}^{\infty} \mathrm{P}(N = n) f^{*n}(t). \tag{3.5}$$

These formulae hold under rather broad conditions which guarantee that the sum converges and the various $f^{*n}(x)$ exist. If one is willing to place some restrictions on the distribution of claim counts N, then (3.5) can be further simplified.

A common approach is to consider the characteristic function (or Fourier transform) of the probability density function of the claim size distribution

$$C[f](t) = \mathrm{E}[\exp(itX)], \tag{3.6}$$

where i is the imaginary unit. Under rather broad regularity and integrability conditions on f, this function exists and is "unique." Thus, if the characteristic function is known then, theoretically at least, the underlying distribution function can be found. A useful property of the characteristic function is that

$$C[f*g](t) = C[f](t)C[g](t) \tag{3.7}$$

if f and g are independent p.d.f.'s. Thus, under conditions sufficient for the sums to exist,

$$C[h](t) = \sum_{n=0}^{\infty} \mathrm{P}(N = n)\, C[f](t)^n. \tag{3.8}$$

If N is further assumed to have a Poisson distribution with mean v, i.e.,

$$\mathrm{P}(N = n) = e^{-v}v^n/n!,$$

then $C[h](t)$ can be written as

$$C[h](t) = \sum_{n=0}^{\infty} e^{-v}v^n C[f](t)^n/n!$$

which reduces to

$$C[h](t) = \exp\{v(C[f](t) - 1)\}. \tag{3.9}$$

Note that $C[h](t)$ is the moment-generating function of the Poisson distribution evaluated at the natural logarithm of the characteristic function of the claim size distribution. Under suitable regularity conditions, this result generalizes to other claim count distributions. That is, the characteristic function of the aggregate distribution is the moment-generating function of the claim count distribution evaluated at the natural logarithm of the characteristic function of the claim size distribution.

Heckman and Meyers [18] present an algorithm which "inverts" this characteristic function. They only require that the probability density function for the distribution of claims by size be a finite step function. Since any (reasonable) probability density function can be approximated as closely as desired by such step functions, conceptually the algorithm they developed should be applicable in any situation.

In addition, they relax the above condition that the claim count distribution be Poisson, with variance and mean equal. Their algorithm also applies to the cases when that distribution is binomial (with variance less than the mean) and negative binomial (with variance greater than the mean). They include a provision for the uncertainty in parameter estimates in the choices of the distributions.

Finally, computer code is provided for the algorithm. The algorithm is computationally rather efficient and can easily be run on a microcomputer with a mathematical co-processor in a reasonable amount of time. In short, Heckman and Meyers provide a very valuable tool to estimate the distribution of T and, for a very wide range of cases, effectively solve that problem.

Another approach to this problem was taken by Panjer [19] and by Sundt and Jewell [20]. In the simplest case, assuming the claim count distribution is Poisson and the p.d.f. of the claim size distribution is discrete and evaluated at equally spaced points, there is a recursive formula which leads to a direct calculation of the distribution of T. Work continues in this area (for example, Willmot [21]).

4. APPLICATIONS IN LOSS RESERVES

It is interesting to note that the majority of the references listed so far either deal with Risk Theory on its own or in relationship to various aspects of ratemaking. There have been some recent papers dealing with risks and uncertainty in loss reserve estimates (see [3], [4], [22], [23], and [24]), but we have been unable to find any which deal directly with considerations which enter with the application of this model to the estimation of variability in loss reserves.

The model of the insurance process provided by Collective Risk Theory seems a natural tool to apply in evaluating the degree of uncertainty in loss reserve estimates. If, for example, under the independence hypotheses listed in Section 2, the distribution of open and IBNR claims (N) is known and the distribution of the size of those claims (X) is also known, the methods outlined in Section 3 all provide ways to estimate the distribution of total reserves (T).

One approach used at this point takes the actuary's best estimate of ultimate claim counts and losses as an estimate of the expected number of claims $E(N)$ and average claim size $E(X)$. Statistical distributions are then selected for each of these quantities.

If the Poisson is chosen as a model of the claim count distribution, then the only parameter to estimate is its mean. Other distributions, such as the binomial and negative binomial, allow for the variance of N to differ from its mean. These are "well behaved" and can be easily accommodated in the algorithm described in [18].

The claim size distribution is usually assumed to be more complex. Common choices include the lognormal, Pareto, and a transformed Gamma, among others. An *ad hoc* approach is to select the distribution to be used, assume that its mean corresponds with the average claim size derived by the actuary's best estimate, and then select the other parameter(s) either judgmentally or based on characteristics of the line under evaluation. This may be all that can be done in situations where data for further analysis is lacking. If sufficient data is available, however, the techniques described by Hogg and Klugman [25] provide powerful tools to select the "proper" distributions.

To better model the distribution of reserves for an insurer or self-insured, accident (report, or policy) years are often considered separately, with separate distributions of claim counts and claim sizes for each year. This has the benefit of preserving differences in relative maturity and maintaining greater homogeneity of claims within each year. The distribution of total reserves can be calculated using convolutions of the distributions for individual years if the various years are assumed to be stochastically independent. The algorithm in [18] allows for such convolution. One "short-cut" sometimes taken is to approximate the 95^{th} percentile, for example, of the distribution of total reserves by the sum of the 95^{th} percentiles of the distributions of reserves for various accident years. A bit of reflection leads to the conclusion that this assumes that the various distributions are perfectly correlated with each other.

There are many possible approaches that can be used to estimate the distributions and resulting reserve variability estimates. What follows here is a discussion of only one possible approach.

This refinement considers the distribution of reserves for an accident year as the combination of the distributions of reserves in three categories: case reserves, development reserves, and IBNR reserves. In this discussion, we consider reserves for reopened claims in the IBNR category. This approach allows closer modeling of the various components of the reserves. These three components also have respectively increasing uncertainty, summarized in the following table:

	Counts	Amounts
Case Reserves	Certain	Certain
Development Reserves	Certain	Uncertain
IBNR Reserves	Uncertain	Uncertain

Distributions for Reported Claim Sizes—One Approach

If we group the first two categories, the case and development reserves, then the statistical uncertainty lies only in the variation of claim sizes, since the number of the claims is known. Given an estimate of the claim size distribution, methods presented in Section 2 could be applied to estimate the distribution of these reserves.

The current distribution of open and reported claims may provide some knowledge of this distribution. For more mature years, one could consider the relationship between the distribution of claims at this stage of development with the "ultimate" distribution of those same claims and incorporate it, with the current distribution, to estimate the ultimate distribution of claims.

As an example of one possible approach, let us assume that the lognormal is an appropriate model for the distribution of X, the claim size random variable. Then $Z = \ln(X)$ has a normal distribution, and the lognormal can be completely parameterized by the mean m and variance s^2 of Z. We select this parameterization for the distribution of X.

It then follows (see, for example, p. 38 of [26]) that maximum likelihood estimators for m and s^2 are obtained from the sample mean and variance of the values $\ln(X_i)$ where X_i are observed claims. As in the normal case, the sample variance, using the number of sample points as the denominator, is a biased estimate for s^2; therefore, a denominator of $n - 1$ is used to estimate s^2.

Suppose, for example, that we are trying to estimate the claim size distribution for open and reported claims for accident year 1981 as of December 31, 1988. That accident year is currently 84 months from the beginning of 1981.

We can calculate the estimators m_{84} and s_{84}^2 of the m and s^2 parameters for reported claims for "mature" accident years at 84 months of development. We can also calculate the estimators m_{ult} and s_{ult}^2 for the distribution of ultimate values of those same claims. Using regression we can find constants which best fit

$$m_{\text{ult}} = a + bm_{84} \textit{ and} \tag{4.1}$$

$$s_{\text{ult}}^2 = c + ds_{84}^2 \tag{4.2}$$

for the "mature" years. These parameters, along with the estimators m_{84}^* and s_{84}^{*2} for the current distribution of claims for accident year 1981 as of December 31, 1988, yield the following estimates of the parameters for the ultimate distribution of currently reported and open claims for accident year 1981:

$$m_{\text{ult}}^* = a + bm_{84}^* \text{ and} \tag{4.3}$$

$$s_{\text{ult}}^{*2} = c + ds_{84}^{*2}. \tag{4.4}$$

Exhibits 1 and 2 provide a numerical example of this approach using purely hypothetical data. In these exhibits, we assume losses for the first seven accident years are sufficiently developed so that we "know" their ultimate distributions and wish to estimate the distribution for accident year 1981.

The distributions of claims reported at 84 months are shown in Exhibit 1. Also shown in Exhibit 1 are the ultimate distributions for the claims reported at 84 months for the first seven accident years, as well as the corresponding parameters from the fitted lognormal distributions.

Exhibit 2 shows the results of the regression and corresponding constants (a and c above) and coefficients (b and d above). Given the lack of significance of the coefficient in the fit for s^2, we assume no relationship between s^{*2}_{84} and s^{*2}_{ult}. We thus use the sample mean and variance for the ultimate distributions for the first seven years as our parameter estimates. The bottom portion of Exhibit 2 then shows the resulting estimates for the parameters of the ultimate distribution for accident year 1981.

At this point, other analyses (e.g., usual reserve estimation techniques) could be used to modify these parameters to reflect the results of those projections. It is a property of the lognormal distribution that the coefficient of variation (ratio of the standard deviation to the mean) can be expressed only in terms of the parameter s^2:

$$c.v.^2 = \exp(s^2)-1. \tag{4.5}$$

Thus, adjustments made to the m^*_{ult} parameter will affect the mean of the final distribution but not its relative variation, as measured by the coefficient of variation. This technique does, however, have the benefit of incorporating information regarding the current distribution of open and reported claims in deriving the estimate for the ultimate distribution of those claims.

We note that there is no chance of zero claims in the lognormal distribution. If we were to use only that distribution as a model for reported claims, then, strictly speaking, the number of claims is not certain, for there may be open claims that will close without payment. This can also be overcome by estimating the portion of those claims which will close with payment separately, possibly also with the use of regression.

Distributions for IBNR Reserves—One Approach

For estimating the distribution of IBNR reserves, both the claim counts and severity are uncertain. The parameters for the claim size distribution could be considered in light of the ultimate value of claims for more "mature" years which were reported after 84 months. The trend in those costs could also be considered in selecting the distribution of claim sizes.

One approach to estimating the distribution of claim counts would be to assume it is Poisson and estimate the expected number of IBNR claims using usual actuarial projection methods. Another approach, similar to that used by Weissner [27], considers the reporting emergence as a statistical distribution with known data truncated from above. Maximum likelihood estimators are then used to estimate the parameters of that distribution. A benefit of this approach is that it can result in estimates of both the mean and variance of the claim count distribution.

This approach begins by postulating a development curve in the form of a probability distribution and then uses maximum likelihood estimators along with known reported claims to estimate the ultimate number of reported claims as well as an approximate distribution of that ultimate. Though the application is in terms of reported claims, there is no inherent reason that the same approach cannot be used to estimate the distribution of ultimate losses directly.

We first assume that the number of claims reported through time t can be expressed as

$$UF(t;\vec{\theta}). \tag{4.6}$$

Here U is the (unknown) ultimate number of claims, and $F(t;\vec{\theta})$ is a cumulative distribution function with parameter(s) $\vec{\theta}$ representing the percent of ultimate claims reported through time t.

In this application, we think of the number of claims reported in time period i as a grouped sample containing f_i points in the interval (c_{i-1}, c_i) from the distribution. We can use methods described in [25] to iteratively approximate the maximum likelihood estimator of the parameter(s) $\vec{\theta}$ given these k observations. To this end, define

$$P_r(\vec{\theta}) = [F(c_r;\vec{\theta})-F(c_{r-1};\vec{\theta})]/F(c_k;\vec{\theta}). \tag{4.7}$$

Here c_{r-1} and c_r are the endpoints of the interval containing the f_r observations. Let f^* denote the total number of claims reported through k time periods, that is,

$$f^* = \sum_{r=1}^{k} f_r. \tag{4.8}$$

Define $A(\vec{\theta})$ to be the matrix composed of the elements

$$a_{ij}(\vec{\theta}) = f^* \sum_{r=1}^{k} \frac{1}{P_r(\theta)} \frac{\partial P_r}{\partial \theta_i}(\vec{\theta}) \frac{\partial P_r}{\partial \theta_j}(\vec{\theta}), \tag{4.9}$$

and let the vector $S(\vec{\theta})$ have the elements

$$S_j(\vec{\theta}) = \sum_{r=1}^{k} f_r \frac{\partial P_r}{\partial \theta_j}(\vec{\theta}). \tag{4.10}$$

With these functions, which involve only first derivatives of the cumulative probability function with respect to its parameters, iteratively calculate

$$\vec{\theta}_m = \vec{\theta}_{m-1} + [A(\vec{\theta}_{m-1})]^{-1} S(\vec{\theta}_{m-1}). \tag{4.11}$$

Now let $h = F(c_k;\vec{\theta}_0)$ be the estimated percentage of claims reported by time c_k. The actual number of claims reported by time c_k can then be thought of as having a binomial distribution with (unknown) mean Uh and variance $Uh(1 - h)$. Assume at this point that the binomial can be approximated by a normal distribution. Thus, approximately,

$$T \sim N(Uh,\ Uh(1 - h)). \tag{4.12}$$

Hence $U = T/h$ is approximately normal:

$$U \sim N(f^*/h,\ f^*(1 - h)/h^2). \tag{4.13}$$

This results in an approximate distribution of IBNR claims I, where

$$I = U - f^* \sim N(f^*/h - f^*,\ f^*(1 - h)/h^2). \tag{4.14}$$

Given these distribution estimates, an estimate of the distribution of IBNR reserves for accident year 1981 as of December 31, 1988 can then be obtained. If it is assumed that this distribution and the distribution of reserves for reported claims are stochastically independent, then an estimate of the distribution of total reserves can be made by convoluting these two distributions.

The assumption of independence may not be too restrictive in this case. As of December 31, 1988, reported and IBNR claims form two distinct populations. It is unlikely that fluctuations in the loss amounts for a fixed number of known claims will lead to fluctuations in the amounts, or counts, of claims yet to be reported. This does not, however, address the question of parameter estimation for these populations and the potential interrelationships there.

As an example of this approach, Exhibit 3 shows a hypothetical claim emergence pattern for the first 84 months of development. We selected a Weibull distribution to model this claims emergence. That distribution's cumulative density function can be written as

$$F(x;\theta_1,\theta_2) = 1 - \exp\{-\exp[\theta_1 \ln(x/\theta_2)]\}. \qquad (4.15)$$

The methods from [25] were then used to derive the parameter estimates shown in Exhibit 3. These parameters result in an h-value of 0.930, with the resulting estimate of the expected number of IBNR claims of 137 with a variance of 147.46.

Combination of Years

The above calculations lead to an estimate of the distribution of total reserves for a single accident year, in this case 1981. Though not explicitly stated, in practice they would probably be calculated for a single coverage or line of insurance. For a multiple line company, however, the distribution of total reserves, for all lines and for all years, is of concern.

If one assumes that the distributions for the various lines of business and accident years are all stochastically independent, the distribution of total reserves could be estimated by convoluting the distributions for individual lines and accident years. In some situations, the assumption of independence may not be too restrictive.

In other situations, however, the reserve distributions for various lines may not be independent; for example, in the bodily injury and property damage portions of automobile liability coverage, some correlations may sometimes be expected, especially in the distributions of the number of claims.

There has been some activity in extending the Collective Risk Model to include such interrelated events. Cummins and Wiltbank in [28] and

[29] consider multivariate models for claim count and size distributions. These models can be thought of as considering the distribution of claims arising from potentially different, but not independent perils. The paper in [28] specifically addresses the automobile liability situation noted above.

5. OTHER AREAS OF UNCERTAINTY

The applications discussed thus far have only addressed one area of uncertainty, the statistical "noise" inherent in the insurance process, *assuming that all distributions are correct*. Not yet addressed are other areas of uncertainty regarding the loss reserve estimates, such as:

1. How close are the selected parameters to the "real" parameters?
2. Are the distributions used in the model correct?
3. Is the Collective Risk Model the right one to use?

None of these questions has been answered yet, nor has the uncertainty they imply been incorporated in the estimated distribution of reserves. The first question, that regarding parameter uncertainty, is sufficiently significant as to be the topic of a paper by Meyers and Schenker [30]. In some situations, the variation due to parameter uncertainty can outweigh the variation from the pure Collective Risk Model itself. Needless to say, this should be recognized in any application of the Collective Risk Model.

Also recognizing the importance of parameter uncertainty, Patrik and John in [13] reserve the term "Collective Risk Model" to a generalization of what we present here. That generalization recognizes parameter uncertainty by considering the parameters themselves as randomly drawn from some probability space.

Often, parameter uncertainty is recognized by "expanding" the variability of the component claim count or size distributions. If data is lacking, such judgmental approaches may be all that is possible.

The possible approaches included above (*"Distributions for Reported Claim Sizes,"* . . . , *"Distributions for IBNR Reserves"*) lend themselves for inclusion of parameter uncertainty. In the claim size distribution estimates for reported claims, the parameter m^*_{ult} is estimated using linear regression. Usual regression theory leads to the conclusion that the variance of m^*_{ult} can be expressed as

$$s_i^2 = (n - 2)SE_i^2/(n - 4), \tag{5.1}$$

where n is the number of points used in estimating the fit, and SE_i is the standard error of the forecast given the observed value for m^*_{84} *("Distributions for Reported Claim Sizes")*.

We now assume that the claim size distribution is lognormal with parameters m^* and s^{*2}_{ult}, where m^* is now unknown, but having a normal distribution with mean m^*_{ult} and variance s_i^2. In this case, the final claim size distribution will again be lognormal with parameters m^*_{ult} and $s^{*2}_{\text{ult}} + s_i^2$. Thus, the uncertainty regarding the scale parameter m^*_{ult} is translated to a widening of the coefficient of variation of the original distribution. Other such "mixings" of distributions can be found in [31].

The bottom portion of Exhibit 2 continues with the example presented above *("Distributions for Reported Claim Sizes")*. For example, for accident year 1981, the standard deviation of the forecast of m^*_{ult} is 0.136 while the fitted s^{*2}_{ult} is 1.921. This results in an adjusted parameter of 1.939 for use with the lognormal distribution.

The maximum likelihood estimator methods presented in [25], as outlined above *("Distributions for IBNR Reserves")*, also provide means to estimate the distribution of those estimators. What follows uses the notation of Section 4 *("Distributions for IBNR Reserves")* and is an application of those methods based on an unpublished presentation made by Gary Venter.

Under suitable restrictions on the cumulative distribution function F, the values of $\vec{\theta}_m$ given in (4.11) converge to the maximum likelihood estimators of the parameters $\vec{\theta}$, call them $\vec{\theta}_0$. Also under suitable conditions, the resulting parameters have a jointly normal distribution with mean $\vec{\theta}_0$ and variance-covariance matrix $[A(\vec{\theta}_0)]^{-1}$.

Now, $h = F(c_k;\vec{\theta}_0)$ is a function of the maximum likelihood estimators $\vec{\theta}_0$ and, following [25, pages 117–118] has an approximate normal distribution with mean $h_0 = F(c_k;\vec{\theta}^*_0)$, where $\vec{\theta}^*_0$ denotes the estimate of the maximum likelihood estimator $\vec{\theta}_0$. Approximately, then,

$$T|h \sim N(Uh_0, Uh_0(1 - h_0)). \tag{5.2}$$

The variance of h can be approximated as

$$\mathrm{Var}(h) = \sum_{i,j=1}^{m} \sigma_{ij}(\vec{\theta}_0^*) \frac{\partial h}{\partial \theta_i} (\vec{\theta}_0^*) \frac{\partial h}{\partial \theta_j} (\vec{\theta}_0^*). \tag{5.3}$$

Here σ_{ij} denotes the i, j element of the approximate covariance matrix $[A(\bar{\theta}_0^*)]^{-1}$. Thus, approximately,

$$T \sim N(Uh_0,\ Uh_0(1 - h_0) + U^2\mathrm{Var}(h)). \tag{5.4}$$

Taking now

$$U_0 = f^*/h_0 \tag{5.5}$$

as an estimate of the expected value of ultimate claims U, then, approximately,

$$U = T/h_0 \sim N(U_0,\ [U_0h_0(1 - h_0) + U_0^2\ \mathrm{Var}(h)]/h_0^2). \tag{5.6}$$

There are admittedly many approximations in this estimation process. It does, however, attempt to directly recognize the variability inherent in the estimate of ultimate claims.

Using these approximations, the distribution of IBNR claims is then approximately

$$I = U - f^* \sim N(U_0 - f^*,\ [U_0h_0(1 - h_0) + U_0^2\ \mathrm{Var}(h)]/h_0^2). \tag{5.7}$$

When compared with formula (4.14), this indicates that parameter uncertainty adds a factor of

$$U_0^2\ \mathrm{Var}(h)/h_0^2 \tag{5.8}$$

to the variance of the original unadjusted distribution. In the example in Exhibit 3, $\mathrm{Var}(h) = 0.000135$ for the fitted values of θ_1 and θ_2. Also shown in Exhibit 3 are the approximate parameter covariance matrix and the partial derivatives used in calculating $\mathrm{Var}(h)$. In this case, the additional variance from (5.8) is 599.01, resulting in an indicated variance in the projection of IBNR claims of 746.47.

In the examples presented here, we have used a single specific method to estimate the parameters of the claim size distribution and distribution of IBNR claims. Both of these methods are stochastic in nature and thus supply information, under certain assumptions, regarding the uncertainty inherent in their particular projections.

Usual actuarial projection methodology as described, for example, by Skurnick [32] or Berquist and Sherman [33] does not begin with an underlying statistical model. Thus, the distribution of the projections does not have a readily apparent statistical form. This problem is compounded in practice where the actuary considers the results of several different projection methods, often yielding different results, and selects a best estimate of what the ultimate losses for a given coverage in a given accident year will be.

As mentioned above, an approach used in these situations is to use the best estimates of ultimate claim counts and severities as estimates of $E(N)$ and $E(X)$ and then to select the claim count and size distributions to have these expected values. Other parameter(s) are then selected to represent the estimated variance in these two distributions and are derived either by considering appropriate distributions of claims or judgmentally. Parameter uncertainty may then be addressed by widening the resulting distributions.

These methodologies do have the strength of addressing different influences which may be apparent in the data. They also allow for the introduction of seasoned judgment in interpreting the results of the projections or influences in the underlying data.

There are also a variety of models which are statistically based. Taylor's work [34] summarizes many different reserve estimation methods, and Ashe [22] provides a discussion of some of the work which has been done to estimate variance in reserve projections using these methods. Of particular note are regression-based methods of Taylor [35] and Kalman Filter-based methods of DeJong and Zehnwirth [36]. Both techniques look only to the historical development of losses for their projections. It could be argued with this data that, put simply, "not all that can happen has happened." If that is true, these methods may end up understating the amount of variation in reserve projections. However, they could be useful to quantify parameter uncertainty in the estimates for the Collective Risk Model as presented here.

The answer to the question of how much uncertainty is added because of the other two questions cited above is not nearly as clear as that for parameter uncertainty. Estimates of parameter variability may address some of the uncertainty inherent in the choice of a particular distribution for the model. This may be further mitigated by reviewing the fits of various distributions to the data available to minimize the chance of picking the "wrong" one from a particular collection. However, it is unlikely that, in actual applications, the second or third questions posed above can be completely answered.

6. CONCLUSION

As can be seen from some of the questions raised, there appears to be more work necessary to completely answer the question "How good are our reserve estimates?" It has been the intent of this paper to present an introduction to Collective Risk Theory for the first time reader, along with a survey of some of the work that has been done which can be used to attempt an answer to this question.

Without proper understanding, many tools can be misused. This is true with Collective Risk Theory. The basic framework only addresses certain portions of the potential variability in reserve estimates. Parameter uncertainty is one significant area not specifically addressed by the basic model; thus, it should be considered in any serious application to quantifying reserve variability. Though some of the techniques outlined here to address parameter uncertainty are necessarily complex and somewhat abbreviated due to the intended scope of this paper, it is hoped that the reader will appreciate the importance of this aspect of the Collective Risk Model.

REFERENCES

[1] "Statement of Principles Regarding Property and Casualty Loss and Loss Adjustment Expense Reserves," Casualty Actuarial Society, May, 1988.

[2] S.P. D'Arcy, "Revisions in Loss Reserving Techniques Necessary to Discount Property-Liability Loss Reserves," *PCAS* LXXIV, 1987.

[3] C.D. Daykin, G.D. Bernstein, S.M. Coutts, E.R.F. Devitt, G.B. Hey, D.I.W. Reynolds, and P.D. Smith, "Assessing the Solvency and Financial Strength of a General Insurance Company," *Journal of the Institute of Actuaries*, 113, 1987, p. 1.

[4] C.D. Daykin, G.D. Bernstein, S.M. Coutts, E.R.F. Devitt, G.B. Hey, D.I.W. Reynolds, and P.D. Smith, "The Solvency of a General Insurance Company in Terms of Emerging Costs," *ASTIN Bulletin*, 17, 1987, p. 1.

[5] R.E. Beard, T. Pentikäinen, E. Pesonen, *Risk Theory, The Stochastic Basis of Insurance*, Third Edition. Chapman and Hall, 1984.

[6] K. Borch, *Mathematical Theory of Insurance*, D.C. Heath & Co., 1970.

[7] H. Bühlmann, *Mathematical Methods of Risk Theory*, Springer-Verlag, 1970.

[8] H.L. Seal, *Stochastic Theory of a Risk Business*, Wiley, 1969.

[9] K. Borch, "Reformulation of Some Problems in the Theory of Risk," *PCAS* XLIX, 1962, p. 109.

[10] T. Pentikäinen, "The Theory of Risk and Some Applications," *Journal of Risk and Insurance* XLVII, 1980, p. 16.

[11] A.L. Mayerson, D.A. Jones, and N. Bowers, Jr., "The Credibility of Pure Premiums," *PCAS* LV, 1968, p. 175.

[12] P. Bratley, B.L. Fox, and L.E. Schrage, *A Guide to Simulation*, Springer-Verlag, 1983.

[13] G.S. Patrik, and R.T. John, "Pricing Excess of Loss Casualty Working Cover Reinsurance Treaties," Casualty Actuarial Society 1980 Discussion Paper Program, p. 399.

[14] G.G. Venter, "Transformed Beta and Gamma Distributions and Aggregate Losses," *PCAS* LXX, 1983, p. 156.

[15] T. Pentikäinen, "Approximative Evaluation of the Distribution Function of Aggregate Claims," *ASTIN Bulletin* 17, 1987, p. 15.

[16] Hon-Shiang Lau, "An Approach for Estimating the Aggregate Loss of an Insurance Portfolio," *Journal of Risk and Insurance* LI, 1984, p. 20.

[17] S. Philbrick, "An Alternative Approach to Estimating Aggregate Loss Distributions," *The Actuarial Review,* v. 14, 1987, no. 3, p. 8.

[18] P.E. Heckman and G.G. Meyers, "The Calculation of Aggregate Loss Distributions from Claim Severity and Claim Count Distributions," *PCAS* LXX, 1983, p. 22.

[19] H. Panjer, "Recursive Evaluation of a Family of Compound Distributions," *ASTIN Bulletin,* 12, 1981, p. 22.

[20] B. Sundt, and W. Jewell, "Further Results on Recursive Evaluation of Compound Distributions," *ASTIN Bulletin,* 12, 1981, p. 27.

[21] G. Willmot, "Mixed Compound Poisson Distributions," *ASTIN Bulletin,* 16, 1986, p. S59.

[22] F. Ashe, "An Essay at Measuring the Variance of Estimates of Outstanding Claim Payments," *ASTIN Bulletin,* 16, 1986, p. S99.

[23] T. Pentikäinen and J. Rantala, "Runoff Risk as a Part of Claims Fluctuation," *ASTIN Bulletin,* 16, 1986, p. 113.

[24] J.N. Stanard, "A Simulation Test of Prediction Errors of Loss Reserve Estimation Techniques," *PCAS* LXXII, 1985, p. 124.

[25] R.V. Hogg and S.A. Klugman, *Loss Distributions,* Wiley, 1984.

[26] J. Aitchison and J.A.C. Brown, *The Lognormal Distribution With Special Reference to its Uses in Economics,* Cambridge University Press, 1969.

[27] E.W. Weissner, "Estimation of the Distribution of Report Lags by the Method of Maximum Likelihood," *PCAS* LXV, 1978, p. 1.

[28] J.D. Cummins and L.J. Wiltbank, "Estimating the Total Claims Distribution Using Multivariate Frequency and Severity Distributions," *Journal of Risk and Insurance* L, 1983, p. 377.

[29] J.D. Cummins and L.J. Wiltbank, "A Multivariate Model of the Total Claims Process," *ASTIN Bulletin* 14, 1984, p. 45.

[30] G.G. Meyers and N. Schenker, "Parameter Uncertainty in the Collective Risk Model," *PCAS* LXX, 1983, p. 11.

[31] G.G. Venter, "Scale Parameter Mixing with Heavy Tailed Loss Distributions," Unpublished, 1985.

[32] D. Skurnick, "A Survey of Loss Reserving Methods," *PCAS* LX, 1973, p. 16.

[33] J.R. Berquist and R.E. Sherman, "Loss Reserve Adequacy Testing: A Comprehensive, Systematic Approach," *PCAS* LXIV, 1977, p. 123.

[34] G.C. Taylor, *Claim Reserving in Non-Life Insurance*, North-Holland, 1986.

[35] G.C. Taylor, "Regression Models in Claims Analysis I: Theory," *PCAS* LXXIV, 1987, p. 354.

[36] P. DeJong and B. Zehnwirth, "Claims Reserving, State-space Models and the Kalman Filter," *Journal of the Institute of Actuaries*, 110, 1982, p. 157.

EXHIBIT 1
SHEET 1

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

ACCIDENT YEAR 1

Claim Size Range	At 84 Months: Number of Claims	At 84 Months: Average Cost	Ultimate: Number of Claims	Ultimate: Average Cost
$0 – $1,000	199	$450	170	$479
1,001 – 5,000	163	2,730	150	2,738
5,001 – 10,000	55	7,366	65	6,866
10,001 – 25,000	48	17,074	63	16,606
25,001 – 50,000	19	36,052	25	37,506
50,001 – 100,000	10	71,898	15	74,917
100,001 – 250,000	5	158,696	9	162,010
250,001 – 500,000	1	369,018	2	341,595
500,001 – 1,000,000	0	—	1	711,158
1,000,001 –	0	—	0	—
Total	500		500	

Parameters of Fitted Lognormal Distributions

	At 84 Months	Ultimate
m	7.396	7.740
s-squared	1.848	1.937

EXHIBIT 1
SHEET 2

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

ACCIDENT YEAR 2

Claim Size Range	At 84 Months		Ultimate	
	Number of Claims	Average Cost	Number of Claims	Average Cost
\$0 – \$1,000	168	\$443	150	\$442
1,001 – 5,000	168	2,477	172	2,585
5,001 – 10,000	65	7,327	62	7,252
10,001 – 25,000	59	15,551	64	17,055
25,001 – 50,000	25	37,613	29	35,638
50,001 – 100,000	14	72,826	18	71,916
100,001 – 250,000	8	170,667	11	160,023
250,001 – 500,000	2	351,781	3	356,221
500,001 – 1,000,000	1	699,609	1	702,665
1,000,001 –	0	—	0	—
Total	510		510	

Parameters of Fitted Lognormal Distributions

m	7.736	7.918
s-squared	1.862	1.880

EXHIBIT 1
SHEET 3

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

ACCIDENT YEAR 3

Claim Size Range	At 84 Months Number of Claims	At 84 Months Average Cost	Ultimate Number of Claims	Ultimate Average Cost
$0 – $1,000	172	$415	166	$445
1,001 – 5,000	167	2,502	173	2,622
5,001 – 10,000	62	7,172	61	7,522
10,001 – 25,000	65	15,775	61	15,177
25,001 – 50,000	27	38,563	31	37,408
50,001 – 100,000	15	74,796	15	65,545
100,001 – 250,000	9	167,488	9	160,523
250,001 – 500,000	2	363,088	3	365,688
500,001 – 1,000,000	1	663,006	1	705,967
1,000,001 –	0	—	0	—
Total	520		520	

Parameters of Fitted Lognormal Distributions

	At 84 Months	Ultimate
m	7.754	7.797
s-squared	1.899	1.896

EXHIBIT 1
SHEET 4

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

ACCIDENT YEAR 4

Claim Size Range	At 84 Months: Number of Claims	At 84 Months: Average Cost	Ultimate: Number of Claims	Ultimate: Average Cost
$0 – $1,000	160	$480	161	$441
1,001 – 5,000	170	2,558	170	2,837
5,001 – 10,000	74	6,886	70	7,456
10,001 – 25,000	65	15,519	67	15,832
25,001 – 50,000	32	36,991	30	35,140
50,001 – 100,000	17	74,283	17	73,015
100,001 – 250,000	9	163,701	11	158,295
250,001 – 500,000	3	370,993	3	345,297
500,001 – 1,000,000	1	720,316	1	702,860
1,000,001 –	0	—	1	2,117,652
Total	531		531	

Parameters of Fitted Lognormal Distributions	At 84 Months	Ultimate
m	7.896	7.897
s-squared	1.862	1.917

EXHIBIT 1
SHEET 5

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

ACCIDENT YEAR 5

	At 84 Months		Ultimate	
Claim Size Range	Number of Claims	Average Cost	Number of Claims	Average Cost
\$0 – \$1,000	151	\$443	140	\$478
1,001 – 5,000	177	2,647	172	2,531
5,001 – 10,000	73	7,809	74	7,858
10,001 – 25,000	71	16,229	73	16,888
25,001 – 50,000	34	35,331	39	32,982
50,001 – 100,000	20	72,039	21	72,711
100,001 – 250,000	11	153,797	15	154,762
250,001 – 500,000	3	363,043	4	335,047
500,001 – 1,000,000	1	703,801	2	679,978
1,000,001 –	0	—	1	1,924,372
Total	541		541	

Parameters of Fitted Lognormal Distributions

	At 84 Months	Ultimate
m	8.003	8.145
s-squared	1.849	1.919

EXHIBIT 1
SHEET 6

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

ACCIDENT YEAR 6

Claim Size Range	At 84 Months Number of Claims	At 84 Months Average Cost	Ultimate Number of Claims	Ultimate Average Cost
\$0 – \$1,000	153	\$436	151	\$450
1,001 – 5,000	181	2,508	170	2,695
5,001 – 10,000	71	7,373	70	7,270
10,001 – 25,000	78	16,035	78	16,929
25,001 – 50,000	35	37,119	37	36,601
50,001 – 100,000	19	77,169	23	68,545
100,001 – 250,000	11	157,721	15	164,521
250,001 – 500,000	3	366,860	5	337,331
500,001 – 1,000,000	1	716,312	2	694,022
1,000,001 –	0	—	1	2,312,174
Total	552		552	
Parameters of Fitted Lognormal Distributions				
m		8.012		8.103
s-squared		1.839		1.975

EXHIBIT 1
SHEET 7

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

ACCIDENT YEAR 7

Claim Size Range	At 84 Months: Number of Claims	At 84 Months: Average Cost	Ultimate: Number of Claims	Ultimate: Average Cost
\$0 – \$1,000	140	\$466	149	\$468
1,001 – 5,000	187	2,697	183	2,587
5,001 – 10,000	72	7,144	72	8,010
10,001 – 25,000	74	15,859	77	16,430
25,001 – 50,000	42	38,555	37	34,613
50,001 – 100,000	26	73,586	23	72,933
100,001 – 250,000	14	158,619	15	156,530
250,001 – 500,000	5	364,353	4	343,411
500,001 – 1,000,000	2	721,218	2	736,468
1,000,001 –	1	2,128,700	1	2,045,068
Total	563		563	

Parameters of Fitted Lognormal Distributions

	At 84 Months	Ultimate
m	8.180	8.105
s-squared	1.909	1.923

EXHIBIT 1
SHEET 8

DISTRIBUTION OF LOSSES FOR CLAIMS REPORTED BY 84 MONTHS OF DEVELOPMENT

	Accident Year 1981	
Claim Size Range	Number of Claims	Average Cost
\$0 – \$1,000	118	\$459
1,001 – 5,000	183	2,707
5,001 – 10,000	84	7,949
10,001 – 25,000	86	17,114
25,001 – 50,000	51	35,679
50,001 – 100,000	26	72,272
100,001 – 250,000	17	151,062
250,001 – 500,000	6	366,299
500,001 – 1,000,000	2	685,736
1,000,001 –	1	2,126,918
Total	456	

Parameters of Fitted Lognormal Distributions

m	8.411
s-squared	1.835

EXHIBIT 2

Regression Estimates of Parameters for Later Accident Years

Accident Year	*m* Parameters At 84 Months	*m* Parameters Ultimate	*s*-squared Parameters At 84 Months	*s*-squared Parameters Ultimate
1	7.396	7.740	1.848	1.937
2	7.736	7.918	1.862	1.880
3	7.754	7.797	1.899	1.896
4	7.896	7.897	1.862	1.917
5	8.003	8.145	1.849	1.919
6	8.012	8.103	1.839	1.975
7	8.180	8.105	1.909	1.923

Regression Results

	m Parameters	*s*-squared Parameters
Constant	3.580	2.905
Coefficient	0.557	−0.527
Standard Error of Coefficient	0.135	0.447
R-Squared	0.773	0.218
Standard Error of Y-Estimate	0.084	0.029

Accident Year	Fitted *m* at 84 Months	Forecast *m*	Standard Deviation of Forecast *m*	Fitted *s*-squared at 84 Months	Forecast *s*-squared	Uncertainty Adjusted *s*-squared
1981	8.411	8.265	0.136	1.835	1.921	1.939

EXHIBIT 3

Example Calculation of Claim Count Expected Value and Approximate Variance

Months of Development	Reported Claims
0 to 12	463
12 to 24	382
24 to 36	369
36 to 48	236
48 to 60	198
60 to 72	100
72 to 84	74
Total Reported	1,822

Fitted Parameters (Weibull)	
Theta(1) =	1.195
Theta(2) =	37.077

Approximate Parameter Covariance Matrix (Inverse of A(Theta))	
0.00145	−0.02309
−0.02309	1.63535

Partial Derivatives of h with respect to	
Theta(1)	0.152
Theta(2)	−0.00601

Var(h) ~ 0.000135

Additional Variance from Parameter Uncertainty
599.01

h = 0.930
$E(U)$ = 1,959
Var(U) ~ 147.46 (Unadjusted for parameter uncertainty)
Var(U) ~ 746.47 (Adjusted for parameter uncertainty)